

# A Dual Sparse Decomposition Method for Image Denoising

Hong Sun

<sup>1</sup>School of Electronic Information  
Wuhan University  
430072 Wuhan, China

<sup>2</sup>Dept. Signal and Image Processing  
Telecom ParisTech

46 rue Barrault, 75013 Paris, France  
Email: hongsun@whu.edu.cn

Chen-guang Liu

<sup>1</sup>School of Electronic Information  
Wuhan University  
430072 Wuhan, China

<sup>2</sup>Dept. Signal and Image Processing  
Telecom ParisTech

46 rue Barrault, 75013 Paris, France  
Email: chenguang.liu@telecomparistech.fr

Cheng-wei Sang

School of Electronic Information  
Wuhan University  
430072 Wuhan, China

Email: sangcw@whu.edu.cn

**Abstract**—This article addresses the image denoising problem in the situations of strong noise. We propose a dual sparse decomposition method. This method makes a sub-dictionary decomposition on the over-complete dictionary in the sparse decomposition. The sub-dictionary decomposition makes use of a novel criterion based on the occurrence frequency of atoms of the over-complete dictionary over the data set. The experimental results demonstrate that the dual-sparse-decomposition method surpasses state-of-art denoising performance in terms of both peak-signal-to-noise ratio and structural-similarity-index-metric, and also at subjective visual quality.

## I. INTRODUCTION

Two main issues are involved in the denoising problem. One is the filtering technique by signal analysis to identify the information underlying the noisy data. The other is grouping technique by clustering technique to provide homogeneous signals for filtering.

Almost all filtering techniques assume that the involved signal should be homogeneous. Therefore, a grouping procedure is generally required before filtering. Many edge detection and image segmentation techniques [1] are used in image denoising. Recently, a nonlocal self-similarity method [2] provides a potential breakthrough for data grouping, which is adopted in this paper.

The filtering technique is developed in the past 50 years or so from many diverse points of view, statistical estimation method, such as Wiener filter, adaptive filter, etc. [3]; transform-domain method, such as Principal Components Analysis [4], wavelet shrinkage [5], etc., and so on. The underlying assumption of these filtering methods is that information in the noisy data has a property of energy concentration in a small linear subspace of the overall space of possible data vectors, whereas additive noise is typically distributed through the larger space isotropically.

However, in many practical cases, some components with low energy might actually be important because they carry information relative to the signal details. On the contrary, when dealing with noise with non-Gaussian statistics, it may happen that some noise components may have higher energies. Consequently, a major difficulty of filtering is to separate

the information details from noise. A way to deal with this problem is cooperative filtering technique, such as Turbo iterative filter [6].

In recent years, sparse coding has attracted significant interest in the field of signal denoising [7] upon an over-complete dictionary. A sparse representation is a signal decomposition on a very small set of components (called atoms) which are adapted to the observational data. The sparse-decomposition based denoising is much better at the trade-off between the preservation of details and the suppression of noise. However, the sparse decomposition is adapted to noisy data so that separating details from noise still is at issue.

In this paper, we propose a dual sparse decomposition method for filtering. The first decomposition is to make an over-complete dictionary to reject some noises which really distributed through the larger space isotropically but to preserve the information details as much as possible. The second decomposition is to identify principal atoms to form a sub-dictionary which preserve well the weak information details and simultaneously suppress strong noises.

This article is organized as follows: Section 2 analyzes some limitations of the classical sparse decomposition for denoising. Section 3 presents the principle of the proposed dual sparse decomposition. Section 4 shows some experimental results and comparisons with state-of-art image denoising methods. Finally, we draw the conclusion in Section 5.

## II. SPARSE DECOMPOSITION FOR DENOISING

We start with a brief description of the classical sparse decomposition and analyze their limitations for denoising.

The sparse decomposition of  $M$  observations  $\{\mathbf{x}_m \in \mathbb{R}^N\}_{m=1}^M$  based on a dictionary  $\mathbf{D} = \{\mathbf{d}_k\}_{k=1}^K \in \mathbb{R}^{N \times K}$ . When  $K > N$ , the dictionary is said over-complete.  $\mathbf{d}_k \in \mathbb{R}^N$  is a basis vector, also called an atom of the dictionary. They are not necessarily independent. With observational data set:

$$\mathbf{X}_{N \times M} = \{\mathbf{x}_m\}_{m=1}^M \quad (1)$$

a dictionary and the coefficients can be the solution of the following equation [8]:

$$\{\mathbf{D}, \boldsymbol{\alpha}_m\} = \underset{\mathbf{D}, \boldsymbol{\alpha}_m}{\operatorname{argmin}} \|\boldsymbol{\alpha}_m\|_0 + \|\mathbf{D}\boldsymbol{\alpha}_m - \mathbf{x}_m\|_2^2 \leq \varepsilon, \quad (2)$$

$$1 \leq m \leq M$$

where  $\|\bullet\|_2$  denotes  $\ell^2$ -norm and  $\|\bullet\|_0$  denotes  $\ell^0$ -norm. In equation (2),  $\boldsymbol{\alpha}_m = [\alpha_m(1) \ \alpha_m(2) \ \dots \ \alpha_m(K)]^T \in \mathbb{R}^{K \times 1}$  is the sparse code of the observation  $\mathbf{x}_m$ . The allowed error tolerance  $\varepsilon$  can be chosen according to the standard deviation of the noise. The sparse decomposition can be written in matrix form as:

$$\mathbf{X}_{N \times M} \approx \mathbf{D}_{N \times K} \mathbf{A}_{K \times M} \quad (3)$$

where the matrix  $\mathbf{A}$  of size  $K \times M$  is composed of  $M$  sparse column vectors  $\boldsymbol{\alpha}_m$ :

$$\mathbf{A}_{K \times M} = [\boldsymbol{\alpha}_1 \cdots \boldsymbol{\alpha}_k \cdots \boldsymbol{\alpha}_M]$$

An estimate of the underlying signal  $\mathbf{S}$  embedded in the observed data set  $\mathbf{X}$  would be:

$$\mathbf{S} = [\mathbf{d}_1 \ \mathbf{d}_2 \cdots \mathbf{d}_k \cdots \mathbf{d}_K] \cdot [\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_m \cdots \boldsymbol{\alpha}_M] \quad (4)$$

The over-complete dictionary  $\mathbf{D}$  in sparse decomposition can effectively capture the information patterns and reject white Gaussian noise patterns. However, we note that the learning algorithm for dictionary  $\mathbf{D}$  by equation (2) would fall into a dilemma of preserving weak details and suppressing noise. On one hand, in order to suppress noise, the allowed error tolerance  $\varepsilon$  in equation (2) should be small enough. As a result, certain weak details would be lost. On the other hand, in order to capture weak details,  $\varepsilon$  cannot be too small. Otherwise some atoms would be so noisy that degrade the denoising performance. Fig. 1 shows an example to show this situation. Taking a noisy image degraded by white noise with standard deviation  $\sigma = 35$  (Fig. 1a), we make two different dictionaries  $\mathbf{D}_H$  and  $\mathbf{D}_L$  (Fig. 1b) by solving equation (2) with  $\varepsilon = 40$  and  $\varepsilon = 35$  respectively. We got two different retrieved images  $\mathbf{S}_H$  and  $\mathbf{S}_L$  (Fig. 1b) respectively by equation (4). Intuitively, the noise is well suppressed in  $\mathbf{S}_H$  but some information details are lost. On the contrary, more details are reserved in  $\mathbf{S}_L$  but it is rather noisy.

Considering the above limitation of sparse-decomposition-based denoising, our idea of dual sparse decomposition is to make a two-stop sparse decomposition: The first step is to make an over-complete dictionary by learning from the observational data with a lower allowed error tolerance  $\varepsilon$  according to equation (2). Thereby, the obtained dictionary  $\mathbf{D}_L$  can capture more information details although it contains some noisy atoms. The second step is to make a sub-dictionary decomposition on  $\mathbf{D}_L$  to reject some atoms too noisy.

### III. SPARSE SUBSPACE DECOMPOSITION

To get the sub-dictionary, we introduce a novel criterion to the sparse subspace decomposition of a learned dictionary and a corresponding index of significance of the atoms.

#### A. Occurrence Frequency of Atom

Atoms  $\{\mathbf{d}_k\}_{k=1}^K$  in the sparse decomposition are prototypes of signal segments. This property allows us to take the atoms as a signal pattern. Thereupon, some important features of the signal pattern could be considered as a criterion to identify significant atoms. We note a common knowledge about the regularity of signal: A signal pattern must occur in meaningful signals with higher frequency even with a lower energy, such as the geometrical regularity of image structures like edges and textures. On the contrary, a noise pattern would hardly be reproduced in observed data even with a higher energy. Therefore, we propose to take the frequency of atoms appeared in the data set as the criterion to identify principal atoms [9]. In fact, the frequency of atoms is a good description of the signal texture [10].

We intend to find out a measurement of the frequency of atom from the sparse codes. Coefficient matrix  $\mathbf{A}$  in the sparse representation (equation (3)) is composed by  $M$  sparse column vectors  $\boldsymbol{\alpha}_m$ . Let us consider the row vectors  $\{\beta_k\}_{k=1}^K$  of coefficient matrix  $\mathbf{A}$ :

$$\mathbf{A} = [\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \ \cdots \ \boldsymbol{\alpha}_M] = \begin{bmatrix} \alpha_1(1) & \alpha_2(1) & \cdots & \alpha_M(1) \\ \alpha_1(2) & \alpha_2(2) & \cdots & \alpha_M(2) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1(K) & \alpha_2(K) & \cdots & \alpha_M(K) \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}$$

where

$$\beta_k = [\alpha_1(k) \ \alpha_2(k) \ \dots \ \alpha_M(k)] \in \mathbb{R}^{1 \times M} \quad (5)$$

Note that the row vector  $\beta_k$  is not necessarily sparse.

Thus, the coefficient matrix  $\mathbf{A}$  can be written by  $K$  row vectors as:

$$\mathbf{A}_{K \times M} = [\beta_1^T \cdots \beta_k^T \cdots \beta_K^T]^T$$

Then equation (4) can be expressed by reordered dictionary and its coefficient as:

$$\mathbf{S}_{N \times M} = [\mathbf{d}_1 \cdots \mathbf{d}_k \cdots \mathbf{d}_K] \cdot [\beta_1^T \cdots \beta_k^T \cdots \beta_K^T]^T \quad (6)$$

Denoting  $\|\beta_k\|_0$  the  $\ell^0$  zero pseudo-norm of  $\beta_k$ , we find that  $\|\beta_k\|_0$  is just the number of occurrences of atom  $\mathbf{d}_k$  over the data set  $\{\mathbf{x}_m\}_{m=1}^M$ . We can define the frequency of the atom  $\mathbf{d}_k$  as  $f_k$ :

$$f_k \triangleq \text{Frequency}(\mathbf{d}_k | \mathbf{X}) = \|\beta_k\|_0 \quad (7)$$

#### B. Subspace decomposition on Over-complete Dictionary

Taking vectors  $\{\beta_k\}_{k=1}^K$  from equation (5), we calculate their  $\ell^0$ -norms  $\{\|\beta_k\|_0\}_{k=1}^K$  and rank them in descending order:

$$\tilde{\mathbf{A}} \triangleq [\beta'_1, \dots, \beta'_k, \dots, \beta'_K] \xleftarrow{\text{sort}} [\beta_1, \dots, \beta_k, \dots, \beta_K] \\ \text{s.t. } \|\beta'_1\|_0 \geq \|\beta'_2\|_0 \geq \dots \geq \|\beta'_K\|_0 \quad (8)$$

Corresponding to the order of  $\{\beta'_k\}_{k=1}^K$ , the reordered dictionary is written as:

$$\mathbf{D} \xrightarrow{\text{sort}} \tilde{\mathbf{D}} = [\mathbf{d}'_1, \dots, \mathbf{d}'_k, \dots, \mathbf{d}'_K] \quad (9)$$

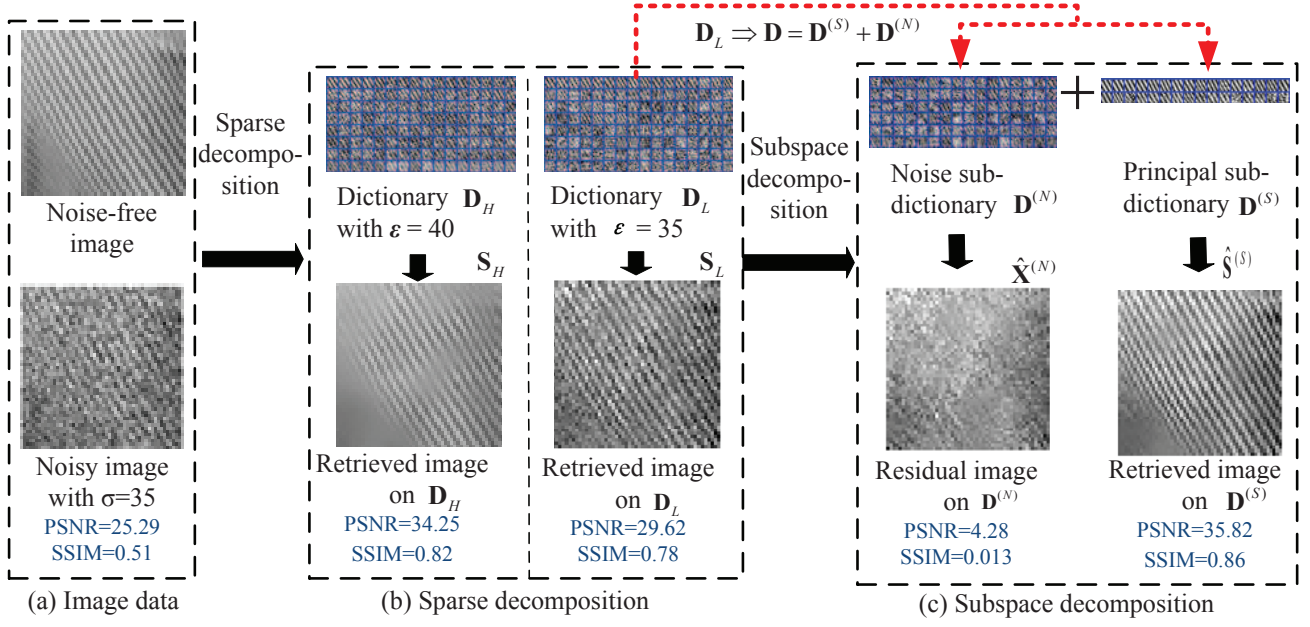


Fig. 1. Principle of Dual Sparse decompositions.

Equation (6) becomes as:

$$\begin{aligned} \mathbf{S}_{N \times M} &= \mathbf{D}_{N \times K} \mathbf{A}_{K \times M} = \tilde{\mathbf{D}}_{N \times K} \tilde{\mathbf{A}}_{K \times M} \\ &= [\mathbf{d}'_1, \dots, \mathbf{d}'_k, \dots, \mathbf{d}'_K] \cdot [\beta'_1{}^T \dots \beta'_k{}^T \dots \beta'_K{}^T]^T \end{aligned} \quad (10)$$

Then, the first  $P$  atoms of  $\tilde{\mathbf{D}}$  can span a principal subspace  $\mathbf{D}_P^{(S)}$  and the remaining atoms span a noise subspace  $\mathbf{D}_{K-P}^{(N)}$  as:

$$\begin{aligned} \mathbf{D}_P^{(S)} &= \text{span}\{\mathbf{d}'_1, \mathbf{d}'_2, \dots, \mathbf{d}'_P\} \\ \mathbf{D}_{K-P}^{(N)} &= \text{span}\{\mathbf{d}'_{P+1}, \mathbf{d}'_{P+2}, \dots, \mathbf{d}'_K\} \end{aligned} \quad (11)$$

In practical application,  $P$  is the threshold of  $f_k$  to separate the principal sub-dictionary from the noise sub-dictionary. We set the maximum point of the histogram of  $\{\|\beta'_k\|_0\}_{k=1}^K$  to  $P$  as:

$$P = \arg \max_k \text{Hist}(\|\beta'_k\|_0) \quad (12)$$

An estimate of the underlying signal  $\hat{\mathbf{S}}$  embedded in the observed data set  $\mathbf{X}$  can be obtained on the principal sub-dictionary  $\mathbf{D}^{(S)}$  simply by linear combination:

$$\begin{aligned} \hat{\mathbf{S}} &= \mathbf{D}_P^{(S)} \cdot \mathbf{A}_P^{(S)} \\ &= [\mathbf{d}'_1, \dots, \mathbf{d}'_k, \dots, \mathbf{d}'_P] \cdot [\beta'_1{}^T \dots \beta'_k{}^T \dots \beta'_P{}^T]^T \end{aligned} \quad (13)$$

Note that  $P \ll K$ .

We show an example of the proposed dual sparse decomposition in Fig. 1(c). The learned over-complete dictionary  $\mathbf{D}$  is decomposed into a principal sub-dictionary  $\mathbf{D}^{(S)}$  and a noise sub-dictionary  $\mathbf{D}^{(N)}$  under the atom's frequency criterion. The retrieved image  $\hat{\mathbf{S}}^{(S)}$  by the dual sparse decomposition method has a super performance at preserving fine details and at suppressing strong noise. We note that the residual

image  $\mathbf{X}^{(N)}$  on the noise sub-dictionary  $\mathbf{D}^{(N)}$  contains some information but very noisy. This is because the atoms of the over-complete dictionary are not independent. The information in the residue image  $\mathbf{X}^{(N)}$  is also in existence in  $\hat{\mathbf{S}}^{(S)}$ .

### C. Application to Filtering

A major difficulty of filtering is to suppress noise Gaussian or non-Gaussian and to preserve information details simultaneously. We use the peak signal-to-noise ratio (PSNR) to assess the noise removal performance:

$$PSNR = 20 \cdot \log_{10} [\max\{\mathbf{S}(i, j)\}] - 10 \cdot \log_{10} [MSE]$$

$$MSE = \frac{1}{IJ} \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} [\mathbf{S}(i, j) - \hat{\mathbf{S}}(i, j)]^2$$

and the structural similarity index metric (SSIM) between denoised image  $\hat{\mathbf{S}}$  and the pure one  $\mathbf{S}$  to evaluate the preserving details performance:

$$SSIM(\mathbf{S}, \hat{\mathbf{S}}) = \frac{(2u_{\mathbf{S}}u_{\hat{\mathbf{S}}} + c_1)(2\sigma_{\mathbf{S}\hat{\mathbf{S}}} + c_2)}{(u_{\mathbf{S}}^2 + u_{\hat{\mathbf{S}}}^2 + c_1)(\sigma_{\mathbf{S}}^2 + \sigma_{\hat{\mathbf{S}}}^2 + c_2)}$$

where  $u_x$  is the average of  $x$ ,  $\sigma_x^2$  is the variance of  $x$ ,  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ , and  $c_1$  and  $c_2$  are small variables to stabilize the division with weak denominator.

From the example shown in Fig. 1, the retrieved image  $\mathbf{S}_H$  actually by the K-SVD filter [11] with the classical sparse decomposition has a high performance with  $PSNR = 34.25$  and  $SSIM = 0.82$  but some information details are obviously lost. On the contrary, the retrieved image  $\mathbf{S}_L$  is noisier with  $PSNR = 29.62$  and  $SSIM = 0.78$  but more information details are reserved. Making a dictionary decomposition on  $\mathbf{D}_L$  noisier but with more details, the retrieved image  $\hat{\mathbf{S}}^{(S)}$  based on the principal sub-dictionary  $\mathbf{D}^{(S)}$  has a higher performance with  $PSNR = 35.82$  and  $SSIM = 0.86$ .



Fig. 2 shows an image filtering result based on the proposed dual sparse decomposition and a comparison with K-SVD algorithm. From the results, the dual sparse decomposition method outperforms K-SVD method by about 1dB in *PSNR* and by about 1% in *SSIM*. In terms of subjective visual quality, we can see that the corner of mouth and the nasolabial fold with weak intensities are much better recovered by the dual sparse decomposition method.

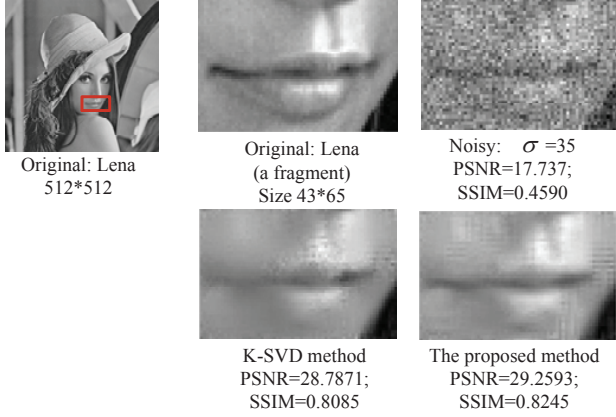


Fig. 2. Image filtering by the dual sparse decomposition comparing with the K-SVD method.

Fig. 3 shows the despeckling results of simulated one-look SAR scenario with a fragment of Barbara image. From the result by a probabilistic patch based (PPB) filter [12] which can cope with non-Gaussian noise, we can see that PPB can well remove speckle noise. However, it also removes low-intensity details. The dual sparse decomposition method shows advantages at preserving fine details and at suppressing strong noise.

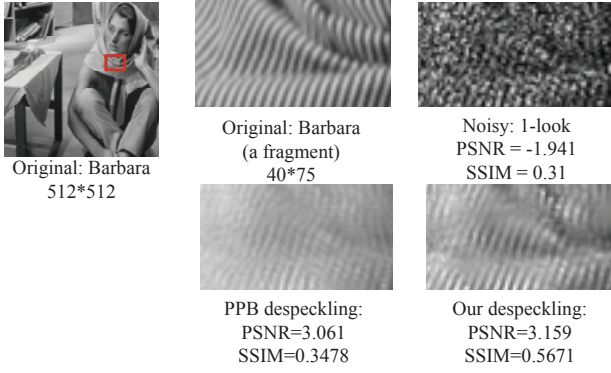


Fig. 3. SAR image despeckling comparing the proposed dual sparse decomposition method with the PPB method.

#### IV. APPLICATION TO DENOISING

In practical applications, our images are generally with spatial complicated scene. On the other hand, the used filtering techniques are generally suitable to homogeneous images.

For image denoising based on the sparse decomposition, the hypotheses of signal sparsity and component reproducibility mean also the condition of homogeneity. In order to make the involved signal homogeneous, we select homogeneous pixels before filtering by a self-similarity measure [2]  $\gamma$ . In applications of image denoising,  $\gamma$  can be specified as Euclidean distance between the reference patch  $\mathbf{x}_i$  and a given patch  $\mathbf{x}_j$  as:

$$\gamma(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_0^2 \quad (14)$$

The smaller  $\gamma$  is, the more similar between  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is. This self-similarity matches well the property of highly repetitive structures of images.

In applications of image despeckling,  $\gamma$  becomes the probabilistic patch-based similarity proposed by [12] as:

$$\gamma(\mathbf{x}_i, \mathbf{x}_j) = (2L - 1) \sum_k \log \sqrt{\frac{\mathbf{y}_i(k)}{\mathbf{y}_j(k)}} + \sqrt{\frac{\mathbf{y}_j(k)}{\mathbf{y}_i(k)}} \quad (15)$$

where  $\mathbf{y}_i = \exp(\mathbf{x}_i)$  and  $L$  the equivalent number of looks.

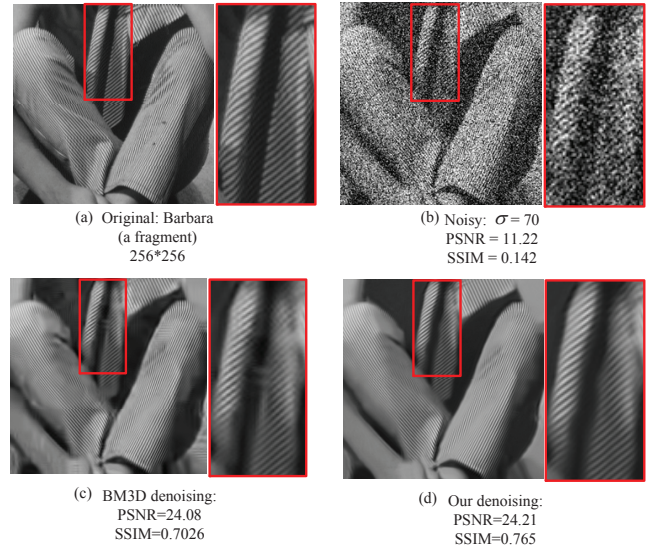


Fig. 4. Denoising for spatial complicated image scene comparing BM3D method.

For a given reference patch  $\mathbf{x}_i$ , we make grouping stacks with its  $\Gamma$ -most similar patches to form a group of data  $\mathbf{z}_i$ . In our experiments, we take  $\Gamma = 90$ . Then we apply a filtering algorithm to each of the data groups  $\mathbf{z}_i, \forall i$ . Our denoising algorithm is presented in Table I:

To compare with the state-of-art denoising algorithm, we take the BM3D method [13], one of the best method nowadays for image denoising. In the BM3D method, a block-matching grouping is also used before filtering. In the experiments, the used dictionaries  $\mathbf{D}$ s are of size are of size  $64 \times 256$  ( $K = 256$  atoms), designed to handle image patches  $\mathbf{x}_m$  of size  $N = 64 = 8 \times 8$  pixels.

Fig. 4 shows the results of denoising an image with a strong additive zero-mean white Gaussian noise and their performances of the dual-sparse-decomposition method and

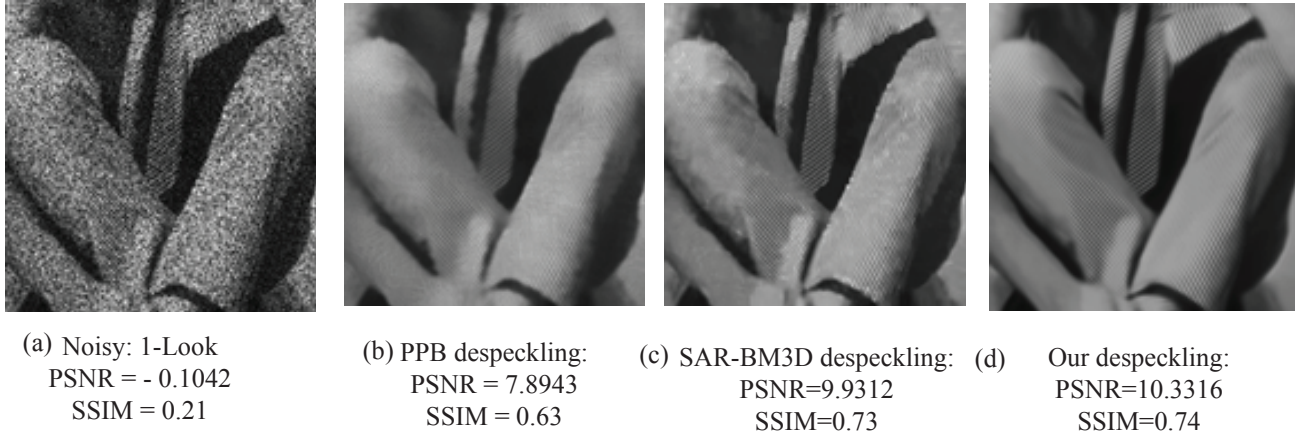


Fig. 5. Principle of Dual Sparse decompositions.

TABLE I  
DENOISING ALGORITHM BASED ON DUAL SPARSE DECOMPOSITION

Input: Image data $\mathbf{X} = \{\mathbf{x}_m\}_{m=1}^M$ (Equ.(1))
Grouping: For patch $\mathbf{x}_i$ , form group $\mathbf{z}_i$ according Eqs. (14) or (15)
Dual sparse decomposition: For each group $\mathbf{z}_i, \forall i$ do
- Sparse decomposition: $\mathbf{z}_i \approx \mathbf{D}\mathbf{A}$ by solving Equ.(2)
- Subspace decomposition: $\mathbf{D} = \mathbf{D}^{(S)} + \mathbf{D}^{(N)}$ by Eqs.(5, (8)-(11)
- Linear reconstruction on $\mathbf{D}^{(S)}$ : $\hat{\mathbf{S}}_i = \mathbf{D}_P^{(S)} \cdot \mathbf{A}_P^{(S)}$ by Eqs. (12)-(13)
Aggregate: to form denoised image <i>weight-average</i> $\{\hat{\mathbf{S}}_i, \forall i\} \Rightarrow \hat{\mathbf{S}}$
Output: Denoised image $\hat{\mathbf{S}}$ .

the BM3D method. Fig. 5 shows the results of despeckling a simulated one-look SAR image with non-Gaussian noise and their performances of the dual-sparse-decomposition method and the SAR-BM3D method [14]. The experimental results demonstrate some advantage of the dual-sparse-decomposition method at preserving fine details and at suppressing speckle noise, also with a better subjective visual quality over the BM3D method.

## V. CONCLUSION

This work present a new signal analysis method by a proposed dual sparse decomposition, leading to state-of-the-art performance for image denoising. The proposed method introduces a sub-dictionary decomposition on an over-complete dictionary learned under a lower allowed-error-tolerance. The principal sub-dictionary is identified under a novel criterion based on the occurrence frequency of atoms. The experimental results have demonstrated that the proposed dual-sparse-decomposition-based denoising method has some advantages both at preserving information details and at suppressing strong noise, as well as provides retrieved image with better subjective visual quality.

It is perfectly possible to straightforward extension the proposed dual-sparse-decomposition to application of feature extraction, inverse problems, or machine learning.

## ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 60872131).

The idea of the dual Sparse decomposition arises through a lot of deep discussions with Professor Henri Maître at Telecom-ParisTech. The mathematical expressions in this paper are corrected by Professor Didier Le Ruyet at CNAM-Paris.

## REFERENCES

- [1] H. Maître, *Le Traitement des Images*. Paris, FRANCE: Lavoisier, 2003.
- [2] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.
- [3] D. G. Manolakis, V. Ingle, and S. Kogon, *Statistical and Adaptive Signal Processing*. New York: McGraw Hill, 2000.
- [4] T. Moon and W. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. New Jersey: Prentice-Hall, 2000.
- [5] D. Donoho, I. M. Johnstone, G. Kerkycharian, and D. Picard, "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society ser. B*, vol. 57, pp. 301–337, 1995.
- [6] H. Sun, H. Maître, and B. Guan, "Turbo image restoration," in *Proc. IEEE International Symposium on Signal Processing and Its Applications (ISSPA2003)*, Paris, France, 2003, pp. 417–420.
- [7] A. Hyvarinen, P. Hoyer, and E. Oja, "Sparse code shrinkage for image denoising," in *Proc. IEEE International Joint Conference on Neural Networks Proceedings*, 1998, pp. 859–864.
- [8] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [9] H. Sun, C. Sang, and C. Liu, "Principal basis analysis in sparse representation," *Science China on Information Sciences*, vol. 60, no. 2, pp. 028 102: 1–3, 2017.
- [10] G. Tartavel, Y. Gousseau, and G. Peyré, "Variational texture synthesis with sparsity and spectrum constraints," *Journal of Mathematical Imaging and Vision*, vol. 52, no. 1, pp. 124–144, 2015.
- [11] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Processing*, vol. 15, pp. 3736–3745, 2006.
- [12] C. A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Processing*, vol. 18, no. 12, pp. 2661–2672, 2009.
- [13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

- [14] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A non-local sar image denoising algorithm based on lmmse wavelet shrinkage," IEEE Trans. Geosci. Remote Sens., vol. 50, no. 2, pp. 606–616, 2012.